Correlation Coeffiecent

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum(x))^2}\sqrt{n\sum(y^2) - (\sum y)^2}} \quad , -1 \leq r \leq 1$$

Ex (1) compute the correlation coeffiecent for the data

x	12	10	14	11	12	9
y	18	17	23	19	20	15

$$r = \frac{(6)(1292) - (68)(112)}{\sqrt{6(786) - (68)^2}\sqrt{6(2128) - (112)^2}} = 0.95$$

x	y	x^2	y^2	xy
12	18			
10	17			
14	23			
11	19			
12	20			
9	15			
68	112	786	2128	1292

* نلاحظ انه r قريبة من الواحد اى انه عند زيادة للتغير الاول يزداد المتغير الثانى بمقدار $\frac{r\, \partial x}{\partial y}$ لانه الارتباط بين المتغيرين ارتباط طردى

* يمكن التنبؤ بقيمة المتغير الثانى الغير موجودة فى الجدول عن طريق

$$y = a + bx \Rightarrow \sum y_i = na + b\sum x_i \longrightarrow ①$$

$$\sum x_i y_i = a\sum x_i + b\sum x_i^2 \longrightarrow ②$$

div ① by ②

$$\frac{\sum y_i}{n} = a + \frac{b\sum x_i}{n} \Rightarrow \mu_y = a + b\mu_x \Rightarrow a = \mu_y - b\mu_x$$

from ②

$$\sum x_i y_i = \left(\frac{\sum y_i}{n} - b\frac{\sum x_i}{n}\right)\sum x_i + b\sum x_i^2$$

$$b\left[\sum x_i^2 - \frac{(\sum x_i)}{n}\right] = \sum x_i y_i - \frac{(\sum y_i)(\sum x_i)}{n}$$

$$b = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \quad —— ③$$

$$a = \mu_y - b\mu_x \rightarrow ④$$

From ③ ⟶ ④ ⟶ ①

Ex: from the following data predict the value of y at $x = 15$

x	12	10	14	11	12	9
y	18	17	23	19	20	15

Sol

$y = a + bx$

$$b = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \quad , \quad a = \frac{\sum y_i}{n} - b\frac{\sum x_i}{n}$$

$$b = \frac{6(1292) - (68)(112)}{6(786) - (68)^2} = 1.47 \quad , \quad a = \frac{112}{6} - (1.47)\frac{68}{6} = 1.91$$

$$y = 1.91 + (1.47)(15) = 23.9$$

\#

"Spearman Rank":

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

اذا كانت البيانات لفظية يمكن استخدام معامل ارتباط "Spearman" بأنه يمثل البيانات المختارة تصاعديا أو تنازليا، وتمثل البيانات الأعلى قيمة 1، والذي تليه 2 وهكذا، واذا تساوى قيمتين نجمع ترتيبهم ونقسمهم على عددهم ونحول البيانات اللفظية الى رقمية ونوجد الفرق بين x و y

Ex: Find Rank of Correlation of the following data:

P = Pass
VG = Very good
G = Good
W = Weak
VW = Very weak
Ex = excellent

1st Term	P	P	VG	G	ex	P	W	VG	G	W
2nd Term	P	G	VG	W	P	G	G	P	W	G

for two esamination of 10 students

Sol.

1st Term	2nd Term	d^2
P (7)	P (7)	$(7-7)^2$
P (7)	G (3.5)	$(7-3.5)^2$
VG (2.5)	VG (1)	$(2.5-1)^2$
G (4.5)	W (9.5)	$(4.5-9.5)^2$
ex (1)	P (7)	$(1-7)^2$
P (7)	G (3.5)	$(7-3.5)^2$
W (9.5)	G (3.5)	$(9.5-3.5)^2$
VG (2.5)	P (7)	$(2.5-7)^2$
G (4.5)	W (9.5)	$(4.5-9.5)^2$
W (9.5)	G (3.5)	$(9.5-3.5)^2$
		205

$$\rho = 1 - \frac{6 * 205}{10(100-1)}$$

$$= -0.24$$

خواص ρ نفس خواص r

$$-1 \leq \rho \leq 1$$

Ex : Find The Rank of correlation ρ of The following data

1st Term	8	3	9	2	7	4	6	1	5
2nd Term	9	5	10	1	8	3	4	2	6

1st term	2nd term	d^2
8 (2)	9 (2)	$(2-9)^2$
3 (7)	5 (5)	$(7-5)^2$
9 (1)	10 (1)	$(1-1)^2$
2 (8)	1 (9)	$(8-9)^2$
7 (3)	8 (3)	$(3-3)^2$
4 (6)	3 (7)	$(6-7)^2$
6 (4)	4 (6)	$(4-6)^2$
1 (9)	2 (8)	$(9-8)^2$
5 (5)	6 (4)	$(5-4)^2$
		12

$$\rho = 1 - \frac{6 * 12}{9(80)} = 0.9$$

Ex: Let $\{(x_i, y_i)\}_{i=1}^{n}$ be ranks individuals distinct Points.

Show That :

$$r = 1 - \frac{6\sum d^2}{n(n^2-1)} = \rho$$

$d_i = x_i - y_i$, $\sum i^2 = \frac{n(n+1)(2n+1)}{6}$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} = \frac{\sum XY}{\sqrt{\sum X^2}\sqrt{\sum Y}}$$

اذا كانت x_i ، y_i نقط مختلفة ، وعددها n فان ترتيبها يكون $1 \to 2 \to n$

Sol.

$X = x - \mu_x$, $Y = y - \mu_y$

The mean of Ranks $= \frac{1+2+3+\cdots+n}{n}$

$\mu_{Rx} = \frac{1}{n}\left[\frac{n}{n}(1+n)\right] = \mu_{Ry}$

مجموع المتوالية = $\frac{N}{2}(\text{ح}_1 + \text{ح}_n)$

$X = x - \mu_{Rx}$

$Y = y - \mu_{Ry}$ $\qquad d = X - Y \Rightarrow \sum d^2 = \sum (X-Y)^2$

$$= \sum X^2 - 2\sum XY + \sum Y^2$$

$$\sum X^2 = \sum (x - \mu_{Rx})^2 = \sum \left(x - \frac{n+1}{2}\right)^2$$

$$= \sum \left(x^2 - (n+1)x + \left(\frac{n+1}{2}\right)^2\right) = \sum (x)^2 - (n+1)\sum x + \frac{(n+1)^2}{4}\sum_{i=1}^{n} 1$$

$$= \sum x^2 - (n+1)\sum x + \frac{n(n+1)^2}{4} \quad \text{—— ①}$$

$$\sum x = 1+2+3+\cdots+n = \frac{n}{2}(1+n)$$

$$\sum x^2 = \frac{n(n+1)(2n+1)}{6} \qquad \text{From ①}$$

$$\sum X^2 = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{2} + \frac{n(n+1)^2}{4}$$

$$= \frac{n(n+1)}{12}\left[4n+2-6n-6+3n+3\right] = \frac{n(n+1)}{12}(n-1)$$

$$\sum X^2 = \frac{n(n^2-1)}{12} = \sum Y^2$$

$$\sum d^2 = \frac{n(n^2-1)}{6} - \sum XY \quad (5)$$

$$\sum XY = \frac{1}{2}\left[\frac{n(n^2-1)}{6} - \sum d^2\right]$$

$$r = \frac{\sum XY}{\sqrt{\sum X^2}\sqrt{\sum Y^2}} = \frac{\frac{1}{2}\left[\frac{n(n^2-1)}{6} - \sum d^2\right]}{\frac{n(n^2-1)}{12}}$$

$$= 1 - \frac{6\sum d^2}{n(n^2-1)} = \rho \ \#$$

Ex: Let μ_x and μ_y be the means of $\{x_i\}_{i=1}^{n}$ and $\{y_i\}_{i=1}^{n}$, $X = x - \mu_x$, $Y = y - \mu_y$ show That

$$r = \frac{\sum XY}{\sqrt{\sum X^2}\sqrt{\sum Y^2}} = \frac{Cov(x,y)}{\sigma_x \sigma_y} \quad \text{where} \quad Cov(x,y) = \frac{\sum XY}{n}$$

التـباين

Sol

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} \quad \text{——} \ ①$$

$$\sum XY = \sum(x-\mu_x)(y-\mu_y) = \sum xy - \mu_y\sum x - \mu_x\sum y + \mu_x\mu_y\sum 1$$

$$\mu_x = \frac{\sum x}{n}, \quad \mu_y = \frac{\sum y}{n}$$

$$\sum XY = \sum xy - \frac{\sum x\sum y}{n} - \frac{\sum x\sum y}{n} + \frac{n\sum x\sum y}{n^2} = \sum xy - \frac{\sum x\sum y}{n}$$

$$\sum X^2 = \sum(x-\mu_x)^2 = n\sigma_x^2$$

$$\sum Y^2 = n\sigma_y^2$$

$$\sigma = \sqrt{\frac{\sum f_i(x-\mu)^2}{\sum f_i}}$$

$$\sigma^2 = \frac{\sum(x-\mu)^2}{n}$$

$$n\sigma^2 = \sum(x-\mu)^2$$

$$r = \frac{\sum xy - (\sum x\sum y)/n}{n\sigma_x\sigma_y}$$

$$\sigma = \sqrt{\frac{\sum f_i x_i^2}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i}\right)^2} \quad \text{where } f_i = 1 \implies \sigma_x = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$\sigma_y = \sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}$$

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{n\sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}\sqrt{\frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2}} \quad \text{Then}$$

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}} \ \#$$

(6)

Ex: show that $\delta^2_{x-y} = \delta^2_x + \delta^2_y - 2r\,\delta_x \delta_y$

الانحراف المعياري الناتج الطرح يمكن الحصول عليه من الانحراف المعياري لـ x و لـ y كل على حده

الحل

$r = \dfrac{\sum X Y}{n\,\delta_x \delta_y}$: $X = x - \mu_x$, $Y = y - \mu_y$

$\delta_x = \sqrt{\dfrac{\sum x^2}{n} - \left(\dfrac{\sum x}{n}\right)^2}$ $\qquad \delta_y = \sqrt{\dfrac{\sum y^2}{n} - \left(\dfrac{\sum y}{n}\right)^2}$

$\delta^2_x = \dfrac{1}{n}\sum [x - \mu_x]^2$ $\qquad \delta^2_y = \dfrac{1}{n}\sum [y - \mu_y]^2$

$\delta^2_{x-y} = \dfrac{1}{n}\sum [(x-y) - \mu_{x-y}]$ $\qquad$ where $\mu_{x-y} = \dfrac{\sum(x-y)}{n}$

$= \dfrac{1}{n}\sum [(x-\mu_x) - (y-\mu_y)]^2$ $\qquad = \dfrac{\sum x}{n} - \dfrac{\sum y}{n}$

$= \dfrac{1}{n}\left[\sum (x-\mu_x)^2 - 2\sum(x-\mu_x)(y-\mu_y) + \sum(y-\mu_y)^2\right.$ $\qquad = \mu_x - \mu_y$

$= \delta^2_x - \dfrac{2n\,r\,\delta_x \delta_y}{n} + \delta^2_y$ **

Ex: From The data

x_i	x_1	x_2	------	x_n
y_i	y_1	y_2	------	y_n

Show That The slop of line of regression is $r\dfrac{\delta_y}{\delta_x}$ where $y = a + bx$, r is correlation coeffiecient

δ_x is standard deviation of x

δ_y " " " " y

الحل

$y = a + bx$ —— ①

$\sum y_i = na + b\sum x_i$ —— ②

$\sum x_i y_i = a\sum x_i + b\sum x_i^2$ —— ③

$\mu_y = a + b\mu_x$ —— ④

$r = \dfrac{\sum X Y}{n\,\delta_x \delta_y}$

$\sum X Y = \sum (x-\mu_x)(y-\mu_y)$ From ①, ④

$= \sum (x-\mu_x)(\cancel{a} + bx - \cancel{a} - b\mu_x)$

$= b\sum (x-\mu_x)^2$

$b\,n\,\delta_x^2 = n\,\delta_x \delta_y\, r$

$b = r\dfrac{\delta_y}{\delta_x}$

b هو ميل الخط $\qquad = b\,n\,\delta^2_x = n\,\delta_x \delta_y\, r$